DOCUMENT RESUME

ED 137 942 95 EA 009 420

AUTHOR Chalupsky, Albert B.; Coles, Gary J.
TITLE The Unfulfilled Promise of Educational Innovation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Apr 77
CONTRACT OEC-0-70-4789
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, N.Y., April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Change; *Educational Innovation; Elementary Secondary Education; Individualized Instruction; Longitudinal Studies; Mathematics Instruction; Program Descriptions; Reading Achievement; *Socioeconomic Influences; *Student Characteristics; Teacher Characteristics; *Time Factors (Learning)
IDENTIFIERS *Longitudinal Study of Educational Practices; *Project LONGSTEP

ABSTRACT

A nationwide, longitudinal study investigated changes in student achievement in grades 1-12 associated with intensive educational innovation. Standardized reading, language, and arithmetic test scores; student and teacher background and attitudinal data; and educational characteristics underlying school programs were obtained over a three-year period. Findings showed a lack of either substantially or consistently positive associations between achievement and program-level emphasis on innovation. Relative to the sample of schools participating in the project, students enrolled in programs with a more moderate emphasis on innovation and individualization showed the greatest improvement. Younger students (those in grades 2 and 3) who achieved substantially more than expected during two consecutive years were exposed to a notably greater amount of class time per day on language arts during the second grade than were the consistent underachievers. The results suggest that educational innovation per se will not necessarily produce dramatic effects on student achievement. Educational quality is not synonymous with innovation. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

30.03

ED137942

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE UNFULFILLED PROMISE OF EDUCATIONAL INNOVATION*

Albert B. Chalupsky

Gary J. Coles

American Institutes for Research
Palo Alto, California

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare (Contract Number OEC-0-70-4789). Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

EA 009 420

*A presentation given at the 1977 annual meeting of the American Educational Research Association, New York City.

THE UNFULFILLED PROMISE OF EDUCATIONAL INNOVATION

Albert B. Chalupsky
Gary J. Coles

American Institutes for Research
Palo Alto, California

Background

During the late 1960s and early 1970s hopes were very high that intensive, innovative educational programs would demonstrate a dramatic impact on student achievement. While these hopes have moderated in recent years, there is still the feeling among many individuals, government officials as well as the general public, that if schools undergo major change they will somehow do much better. As a result, many efforts continue to be sponsored at the federal and state levels to encourage such change, either through the development of new programs or through increased emphasis on the dissemination/diffusion of innovations by educational "change agents." The wholesale adoption of educational innovations, however, is an expensive process, and there is no guarantee that the added costs will necessarily be justified by improved program effectiveness.

Educators and noneducators alike have shown a growing awareness of the lack of--and need for--evidence as to whether or not innovative educational practices are indeed better than the more traditional approaches. In response to this need, the U. S. Office of Education in 1969 awarded a contract to the American Institutes for Research to develop a design for a study of the overall effectiveness of highly intensive, innovative educational practices on students in grades 1 through 12. The initial design was developed, implemented on a limited basis, and modified during the 1969-70 school year. Full implementation began during the 1970-71 school year.

Project Objectives

The specific objectives of Project LONGSTEP (the Longitudinal Study of Educational Practices) were as follows:

- to design a system for studying the educational environments of students in terms of meaningful dimensions which underlie various educational approaches or "treatments";

- to establish a large-scale data base consisting of treatment characteristics, student characteristics, and educational outcomes for a select sample of educational programs involving intensive and highly innovative educational practices;
- to determine, as comprehensively as possible over a three-year period of time, the impact of intensive innovation upon student performance on standardized achievement tests and on measures of educationally relevant attitudes; and
- to attempt to identify the dimensions of educational components (present in a select sample of highly intensive, innovative programs) that exhibit the greatest impact on student outcomes.

For purposes of this study, "intensive innovation" meant the implementation of a new program encompassing a significant proportion of students, entailing a major alteration of school procedures, and involving a high investment of resources.

Study Approach

The general emphasis of Project LONGSTEP was on the identification of changes in student achievement that occur as a result of intensive educational innovation. Highlights of the methods used to identify and document innovations and to assess student achievement are discussed next.

Selection of Schools and Students

During the early stages of the study, a nationwide search was undertaken to locate educational programs incorporating intensive, innovative practices. Guidelines for program selection emphasized program scope and intensity, instructional content, anticipated program continuation, and willingness to cooperate in a multiyear study.

Through a combination of extensive literature search, interviews with program staff, outside consultant review, and site visiting, schools located in 14 school districts were selected and agreed to participate. Of these 14 districts, 13 participated throughout the entire implementation phase and provided the data on which this report is based. Schools and districts were located in nine states: California, Florida, Kentucky, Minnesota, Pennsylvania, Texas, Utah, Washington and Wisconsin. Altogether,

some 30,000 students, 80 schools, and 1,500 teachers participated in the study during its three years of implementation.

At the time of this study, most of the participating educational programs were supported largely by local funding. However, many had their roots in educational research and development activities at least partially funded by the federal government.

Sample Description

The 13 districts participating in Project LONGSTEP provided variability along various dimensions, as illustrated by a number of school and community characteristics noted during the 1970-71 school year. The communities served by the school districts ranged from 2,500 to over 600,000 in population and varied from rural to urban-metropolitan in setting. Their diversity in socioeconomic level is notable, as evidenced by the percentage of students in the participating schools receiving free or reduced-price lunches. This number ranged from less than 1 percent to nearly 30 percent. The reported instructional cost per pupil in the participating school districts varied from a low of $540 to a high of $1,050. Another indicator of the diversity in the sample was the percentage of minority group students in the participating schools; this ranged from less than 1 percent to over 30 percent.

The educational innovations encompassed by Project LONGSTEP were those that were of particular concern during the late 1960s and early 1970s--most are still of concern today. These innovations included team teaching, multimedia emphasis, unique school design, use of paraprofessionals, variations in scheduling, and teacher-developed materials, as well as independent study, student selection of materials, and a number of other practices typically associated with individualized instruction. An overview of the major educational features present in our sample at the time programs were selected, as well as the distribution of such features across districts, is presented in Table 1.

Caution should be exercised in drawing conclusions from this table, however, since the same label may have encompassed significantly different activity elements from one school district to another. By the same token, different labels were often applied to very similar activities. Further,

TABLE 1

Overview of Major Educational Features Occurring in One or More Schools of the Districts Participating in Project LONGSTEP

Major Educational Features	School Districts												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Small-group activities	X	X	X	X				X	X	X		X	X
Team teaching	X	X	X	X	X	X	X	X	X				
Multimedia emphasis	X	X	X	X	X	X	X	X	X	X		X	
Ungraded curriculum		X	X	X	X		X	X	X				
Large-group instruction		X	X					X	X	X	X	X	X
Independent study	X	X	X	X	X			X		X		X	X
Unique school design		X	X				X	X					
Differentiated staffing	X	X				X							
Volunteer aides	X	X	X	X	X	X	X	X	X	X			
IPI reading/math		X			X	X			X				
Project PLAN							X		X				
Individualized instruction (other than PLAN or IPI)	X	X		X		X	X	X	X	X		X	X
Student selection of materials	X	X	X			X		X	X	X			
Modular scheduling	X	X	X	X				X					
Teacher sensitivity training						X							
Intensive guidance and counseling						X							
Home visitation						X							
Traditional/conventional classrooms	X		X	X	X	X	X	X	X	X	X	X	X
Teacher-developed curriculum materials	X	X	X	X		X		X	X	X	X	X	X

it should be stressed that the educational activities listed in the table were often given these designations by the school systems. They did not necessarily reflect the underlying educational processes that were the primary concern of this study. Nevertheless, Table 1 does provide a gross indication of the wide variation that existed in the Project LONGSTEP sample.

Data Collection Instruments and Schedule

The data collection instruments used in this study provided information on student cognitive performance, student characteristics, educational experiences, and teacher characteristics. Student cognitive achievement was measured by either of two commercially developed instruments, the Comprehensive Tests of Basic Skills (CTBS) and the California Short Form Test of Mental Maturity (CTMM), 1963 revision. The CTBS was used for spring testing in grades 2-12 and the CTMM for testing in grade 1. Student and teacher background characteristics and attitudes were assessed by questionnaires developed specifically for Project LONGSTEP.

In order to investigate the relationships between educational practices and educational outcomes in a very diverse group of schools and programs, a system was needed for describing and quantifying the educational experiences that each student had during a school year. It became evident early in the development of this system that, although there are a limited number of labels used by school districts to describe their educational activities, in practice there is a wide disparity among educational approaches sharing the same descriptive label. For example, "modular scheduling" is a commonly-used term for what are in fact various configurations of class schedules. At the same time, programs bearing different labels often turn out to be very much alike. Thus, one district's program of "individualized instruction" could be much the same as another district's "nongraded" program. Therefore, an Educational Experience Analysis Guide was developed so that complex educational experiences could be described and quantified with respect to specific observable characteristics (rather than on the basis of variously defined local labels). AIR staff used this guide to document the basic educational attributes of the school programs in which participating students were enrolled. Information on educational

experiences was gathered from interviews with principals and teachers, from classroom observations, and from existing school documentation.

The guide made it possible to locate the educational experiences of participating students on a continuum ranging from traditional to innovative. Use of the guide also ensured that all students in a school who were receiving essentially the same basic educational experiences would be identified as belonging to the same "educational experience group" and be distinguished from students receiving different educational experiences, even though both groups of students may have been participating in the same school "program."

During the 1970-71 school year, this documentation methodology led to the identification of 141 educational experience groups; during the 1971-72 and 1972-73 school years, respectively, 167 and 228 groups were isolated. These groups represented all the different kinds of basic educational experiences provided by the schools for students participating in Project LONGSTEP. An illustration of the diversity of our sample is shown by the characteristics of the five school districts that contained language arts groups for the third grade during the 1972-73 school year. Figure 1 shows the average scores for these five districts on ten key educational indices (derived from items on the Educational Experience Analysis Guide), plus student socioeconomic status, and the number of minutes per day of class time spent on language arts activities.

One final data collection guide, a classroom observation instrument, was developed to document classroom characteristics such as physical environment, study arrangements, and access to resources, as well as teacher and student activities such as degree of grouping, focus of activities, and use of materials.

In order to collect all the necessary data on students, teachers and educational experiences, AIR staff members typically visited each site three times during a school year. Site visits generally ranged from five to eight days in length. Documentation of educational experiences was usually accomplished during the initial visit each year.

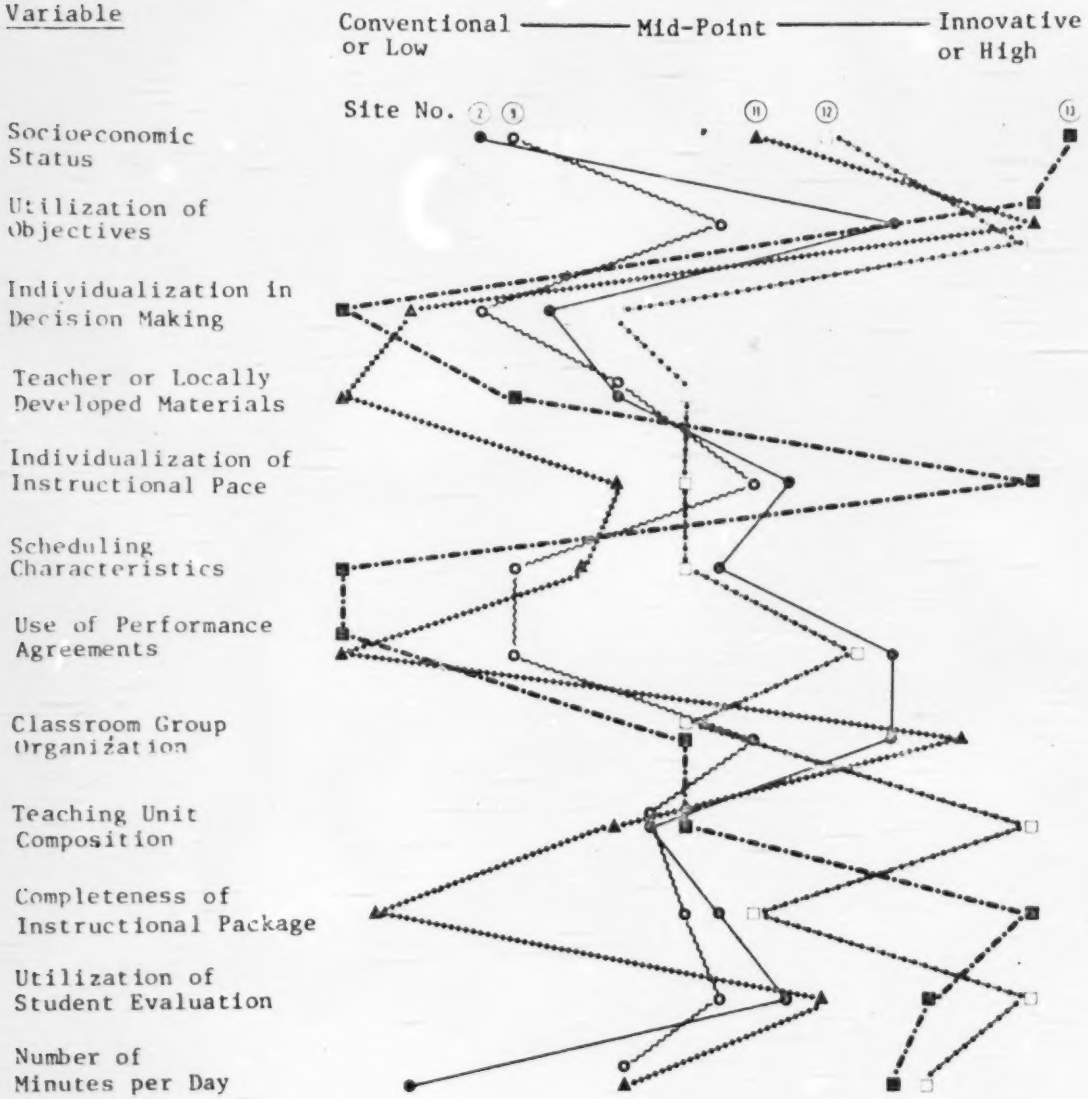

Figure 1. Profiles for language arts groups for the five districts/ sites containing third grade students during the 1972-73 school year.

Overview of the Data Base

Project LONGSTEP represents a pioneering effort in educational evaluation. It is perhaps the first large-scale, longitudinal study to associate the educational attributes underlying school programs and the characteristics of teachers with only those individual students who were directly exposed to their influence. The vast amount of data collected is also notable--other studies may have gathered information on more students, but for sheer quantity and complexity of data, this study is unique. When all the tests and questionnaires administered to students and teachers and all the Educational Experience Analysis Guide data collected on language arts, mathematics, social studies and science treatments over a three-year period are considered, some 3,500 individual items of data could be associated with a single student. When the scales or indices derived from all instruments are also considered, the total number of variables could increase to approximately 4,000 per student.

Because of the potential of such a large and detailed data base, a series of generalized data files and codebooks were produced for the U.S. Office of Education so that these data can be made available to other researchers.

Analysis Approach

While Project LONGSTEP involved students in grades 1-12, the analyses reported here involved only three groups of students: those who started out as first graders in 1970-71, those who started out as fourth graders in 1970-71, and those who started out as sixth graders in 1970-71. These grade levels were selected for initial analysis primarily because their data would permit the examination of educational growth occurring during the primary years (grades 1 to 3), the elementary years (grades 4 to 6), and the transition years from elementary to middle or junior high school (grades 6 to 8). A detailed presentation of the methods and findings is contained in Volume I of the Project LONGSTEP final report (Coles, G. J. Chalupsky, A. B., Everett, B. E., Shaycoft, M. F., Rodabaugh, B. J., and Danoff, M. N., 1976).

The measures that were related to posttest achievement included each student's sociocconomic status (SES, and pretest score; the educational attributes to which each student was exposed, including the overall amount of innovation and individualization in the school program; the amount of class time allocated by that program to language arts or mathematics instruction; and the qualifications of each student's teachers. Amount of innovation was measured by an index called Level of Innovation. This scale was equal to the sum of the following scales derived from the Educational Experience Analysis Guide: Individualization in Decision Making, Individualization of Instructional Pace, Use of Performance Agreements, Utilization of Student Evaluation, Utilization of Objectives, Teacher or Locally Developed Materials, Scheduling Characteristics, Classroom Group Organization, Teaching Unit Composition, and Completeness of Instructional Package. Our measure of program emphasis on individualized instruction, Degree of Individualization, was defined as the sum of the first four of these scales.

Although Project LONGSTEP collected data for three consecutive school years, two pretest/posttest analyses (Year 1 versus Year 2 and Year 2 versus Year 3) were viewed as the most powerful approach for the initial analysis of these data. Because of the multilevel nature of the study's data base, project staff co cluded that overly complex analytic models should not be utilized until the data and the relationships among the basic variables were more clearly understood.

In order to determine whether intensive innovation was related to substantial gains in achievement in Project LONGSTEP's diverse sample of schools, the study assumed that if innovation does have a dramatic impact on achievement, then:

- variation with respect to intensity of innovation, even within a sample of generally innovative schools, should be positively correlated with achievement;
- such impact should be a general trend across different classrooms, schools and school districts and should be consistently present in consecutive school years;

- the achievement gains demonstrated by Project LONGSTEP's generally innovative sample of schools should noticeably exceed the gains shown by national norms; and
- such impact should be large enough to be demonstrated by a number of different methods, including those involving the kinds of statistical adjustments normally required in analyzing nonexperimental data.

Because Project LONGSTEP collected data from students in ongoing school programs, it was deemed essential that the study's key hypotheses be evaluated from a number of slightly different methodological perspectives so as to minimize the possibility that findings would be highly method- or assumption-dependent. Conclusions could then be based on converging lines of evidence. For this reason, the analyses were, by design, as intensive as possible within project time and cost constraints.

Findings in Review

So that they might be summarized in a concise and meaningful way, the Project LONGSTEP findings have been organized into four sections. Each section is headed by the overall, basic question to which the findings pertain.

Were Students of Different Socioeconomic Status and Pretest Levels Exposed to Different Kinds of Treatments?

In a number of grades, lower SES and lower pretest students were members of schools and instructional programs that spent more time on language arts and mathematics activities. There also was a very slight tendency across all analyses for lower SES students to be exposed to less innovative and less individualized programs of instruction. Even the most notable of these trends, however, was extremely small relative to the magnitude of the differences in socioeconomic status among the groups studied. The general finding that student SES and pretest background were not both highly and systematically related to the characteristics of educational approaches means that it is not likely that such factors biased (in a constant manner across all analyses) the overall results of Project LONGSTEP concerning the impact of intensity of innovation.

To What Extent Is Achievement Related to Pretest and Socioeconomic Status?

Both initial achievement level (i.e., pretest) and student socioeconomic status (SES) are usually substantially related to posttest achievement in school effects studies. The results of Project LONGSTEP are no exception. It is notable, however, that:

- pretest was a vastly better predictor of posttest achievement than was SES;
- the predictive value of SES was reduced almost to zero after the predictive overlap with pretest was taken into account; and,
- the relationship between pretest and posttest was still very large even after student differences with respect to SES and after group differences with respect to mean pretest level were taken into account.

While not the major purpose of the study, these findings have implications for both educators and researchers. Based on a diverse sample of schools, such findings suggest that when pretest is known, it is not necessary to use socioeconomic status as a predictor of individual student posttest achievement. They also imply that SES cannot be expected to serve as a reasonably accurate substitute measure of initial achievement (i.e., pretest) level.

Did Intensive Educational Innovation Result in Substantial Growth in Achievement?

Project LONGSTEP findings with respect to the impact of educational innovation can be organized into two general categories. The first group of findings evaluated the overall amount of educational growth demonstrated by each of the samples analyzed and compared this growth with that of the CTBS norm sample. The second group of findings related variation in intensity of innovation to variation in achievement growth.

Growth in achievement and norm comparisons. Comparisons among average pretest performance, average posttest performance, and national norms showed that:

- the greatest gains in reading, language and arithmetic achievement occurred in the earlier grades;

- the magnitude of yearly gains in achievement tended to decrease with each higher grade;
- for all grades, achievement gains seemed to be slightly larger in arithmetic than in language arts (reading and language);
- although gains were not dramatic, achievement growth during third grade was somewhat greater than that in the norm sample; and
- Project LONGSTEP students, on the average, did not do conspicuously better on posttest (relative to national norms) than they did on pretest.

In summary, the expectation that substantial yearly gains in student achievement would occur for a sample of intensive, innovative educational programs is not supported by these findings.

Growth in achievement and innovative emphasis. The most acceptable approach to analyzing data obtained by nonexperimental/associational studies (like Project LONGSTEP) is to use a number of different analysis methods and look for consistent results across methods. Therefore, Project LONGSTEP analyzed:

- overall differences in achievement growth among educational approaches;
- an educational growth model in which achievement was related to innovative emphasis, number of minutes of instruction per day, pretest, socioeconomic status (SES), and teaching qualifications;
- a posttest achievement score statistically adjusted for pretest and SES differences;
- groups of students with similar educational experiences who, on the average, performed either much better or much worse than was expected from their pretest and SES; and
- consistently overachieving and underachieving students.

The major findings based on these methods are reported next with respect to reading and arithmetic achievement. No separate discussion is presented for language achievement since these findings closely paralleled those for reading. A later section contains a more general summary of project findings.

Reading achievement. The analyses of reading achievement indicated that:

- meaningful differences in educational growth did exist among different educational experience groups, especially those involving students in the early elementary grades;
- differences among educational experience groups decreased with each higher grade;
- in spite of group differences in achievement, no consistent overall relationship existed between innovative intensity (as indexed by the study's measure of general innovativeness, Level of Innovation) and posttest reading performance, or between individualization emphasis (as indexed by Degree of Individualization) and posttest reading performance;
- in the only series of analyses in which growth in reading achievement was related to overall innovation and individualization to a substantial degree (during the third grade), the impact of Level of Innovation (or Degree of Individualization) was negative--that is, the greatest educational growth occurred in programs with a more moderate emphasis on innovation;
- the study's measure of teacher experience, Teaching Qualifications, was not related to posttest performance to any meaningful degree in our analyses; and
- the analysis of unusually effective reading groups provided no clear evidence that Level of Innovation or Degree of Individualization was consistently related to achievement. This was also true of the analysis of students who demonstrated significant overachievement or underachievement during two consecutive school years. In fact, the consistent overachievers actually tended to be members of programs with a lower Level of Innovation or Degree of Individualization.

An additional finding regarding the study's measure of amount of schooling, number of minutes per day of class time on language arts, is noteworthy. Younger students (i.e., those in grades 2 and 3) who exhibited

unusually large gains in reading achievement during two consecutive school years were exposed to much more language arts instruction as second graders than were students with a notable lack of growth (i.e., 113 minutes per day compared with 85 minutes). The consistent overachievers, as third graders, were exposed to about 20 minutes of instruction per day less than during the previous school year, and yet they again demonstrated larger than expected gains in achievement. These results suggest that increased exposure to language arts during the second grade may significantly improve the chances of some students to demonstrate substantial gains in reading achievement, even in later grades.

Arithmetic/Mathematics achievement. Meaningful differences among educational experience groups existed with respect to arithmetic posttest achievement, especially at the lower grade levels. In fact, differences among these groups decreased with each higher grade. Other analyses showed that:

- Level of Innovation or Degree of Individualization was negatively, rather than positively, related to growth in arithmetic achievement. This appeared to be the most notable result in the samples as a whole. Therefore, the hypothesis that greater overall emphasis on innovation or individualization is positively and consistently related to student achievement in arithmetic was clearly not supported.

- Neither Level of Innovation nor Degree of Individualization demonstrated a dramatic or consistently positive effect on student arithmetic achievement in (1) the analyses of unusually effective educational experience groups and (2) the analyses of students who demonstrated significant overachievement or underachievement for two consecutive school years. In fact, relative to underachieving students, the students who were identified as overachievers, on the average, were members of programs with a more moderate emphasis on innovation or individualization.

Overview of achievement and innovative emphasis. In summary, no evidence could be found that either of the major treatment variables--Level of Innovation and Degree of Individualization--was substantially and positively correlated with posttest performance. Further, preliminary analyses showed that, on the average, these findings applied equally well to students at different SES or pretest levels.

Meaningful differences in the teacher qualifications of groups performing at a higher or lower level than predicted were infrequent. Whenever they did occur, however, better teaching qualifications (as indicated by such items as education, teaching experience, certification and type of appointment) were associated with better student performance. And finally, younger students (i.e., those in grades 2 and 3) who achieved substantially more than expected during two consecutive years were exposed to a notably greater amount of class time per day on language arts during the second grade than were the consistent underachievers. Even though they spent considerably less class time per day on language arts in grade 3, the overachievers again demonstrated dramatic gains.

Can Schools Have a Substantial Impact on Educational Achievement?

The findings show that substantial educational growth was not uniquely associated with innovative emphasis. Even small amounts of growth were not positively and consistently related to our measures of innovation, Level of Innovation and Degree of Individualization. It should not be concluded, however, that substantial growth was absent in the schools and students participating in Project LONGSTEP. On the contrary, as discussed previously, a number of educational experience groups were identified whose students, on the average, did much better than was predicted from their pretest and SES level. For example, as indicated in Figure 2, the average achievement gains of sixth grade students in the most effective reading treatment groups differed dramatically from those exhibited by students in the groups that did much worse than was predicted from their pretest score and their socio-economic status. The students in the poorer performing reading groups shown in Figure 2 had average pretest scores ranging between the 56th and 74th percentiles; their average posttest scores dropped slightly, ranging between the 41st and 66th percentiles. In contrast, the sixth grade

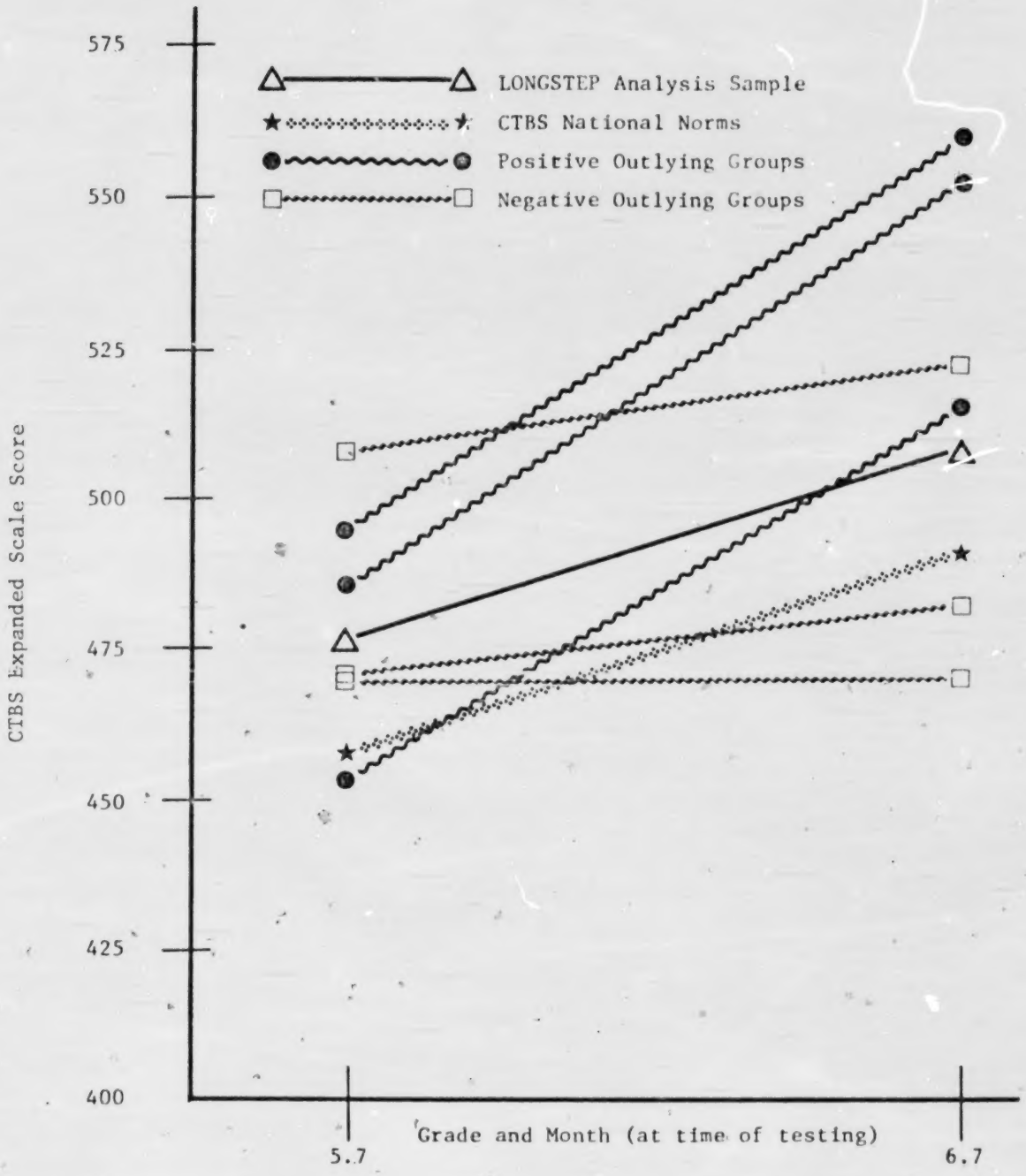

Figure 2. Educational experience groups where students achieved more or less than was expected from their pretest and SES level (Spring 1973 Reading Analysis).

students in the high performing reading groups, who had average pretest scores ranging between the 48th and 69th percentiles, improved noticeably, with average posttest scores ranging between the 63rd and 79th percentiles.

Other examples of dramatic gains in achievement were shown in the analyses identifying individual students who, for two consecutive years, demonstrated posttest achievement differing greatly from that expected on the basis of their pretest and SES scores.

The analyses of unusually effective groups and the analyses of students with unusual growth patterns both provide evidence that large gains in cognitive achievement were occurring in Project LONGSTEP. Thus, even though the overall project findings showed that dramatic school effects were not associated with intensity of innovation, the findings noted here suggest that substantial achievement gains, over and above those expected on the basis of pretest and SES, do occur in public schools. It should also be remembered that the overall analyses showed that small but meaningful posttest differences existed among our educational experience groups, although they were not uniquely associated with the project's measures of innovative intensity.

These findings certainly are not surprising to teachers, principals, parents or even researchers who have personally witnessed impressive growth in cognitive skills. They are, however, somewhat more "newsworthy" in terms of educational evaluation results. Taken together, then, these findings must lead us to conclude that:

- different educational approaches do produce meaningful and important differences in achievement (especially in the early elementary grades); and
- although substantial gains in achievement were not demonstrated by a majority of Project LONGSTEP programs/schools and students, unquestionably dramatic improvements in reading, language and arithmetic skills were found in some schools.

Future research and evaluation efforts should be directed at identifying the determinants of such significant educational gains.

Overall Conclusions and Implications

The single most important and well-documented finding of Project LONGSTEP was the lack of either substantial or consistent association between student achievement and overall Level of Innovation across grades. In fact, relative to the sample of schools participating in the project, students enrolled in programs with a more moderate emphasis on innovation and individualization showed the greatest improvement.

From the standpoint of time spent per day on language arts activities, it appeared that consistently overachieving second and third graders in Project LONGSTEP typically were members of programs that allocated more time to such instruction in the second grade. These findings suggest that increasing the amount of class time per day, especially in the first and second grades, may be a worthwhile strategy for improving student performance in language arts.

Our various analyses also suggest that the impact of the educational practices studied was inversely related to grade level. Specifically, it appeared that differences among educational experience groups, independent of the effects of student SES level, initial achievement level, and Teacher Age, SES Background and Teaching Qualifications, were greatest for second and third graders, less for fifth and sixth graders, and almost nonexistent for seventh and eighth graders. These findings tend to support the view that the greatest potential payoff from the investment of future educational research and instructional development funds may come from increased attention to improving education in the early elementary grades. It is suggested that further effort be expended to gain a better understanding of the educational practices that show the greatest benefit at these early grades. The results of such research could then serve as the basis for designing educational strategies that could have a relatively greater overall impact on American education.

It is less obvious why one particular aspect of much recent innovation, that is, increased individualization of education, has not been shown to be clearly and consistently effective in Project LONGSTEP. Some possible explanations for this finding will be discussed, but it should be emphasized that these are purely speculative. One possibility is that not

all modes of "individualization" operate effectively and that our index, Degree of Individualization, failed to tap some crucial differences between individualized programs that are effective and those that are not. The program characteristics measured by this index included four scales: Individualization in Decision Making (covering decisions on topic selection, materials selection, and sequencing decisions); Individualization of Instructional Pace; Use of Performance Agreements; and Utilization of Student Evaluation (covering the extent to which tests are used for evaluating student needs, and whether or not such evaluation serves as a basis for modifying a student's instructional program).

The amount of independence and decision making left to the student could also be crucial. While tailoring the educational process or teaching strategy to an individual student's needs or learning style may be very effective, allowing the student a great deal of independence in selecting the topics to be studied, in deciding the amount of time to be spent on a particular topic, and in establishing the level or standard of attainment during a school year could have a negative impact on performance (as measured by standardized achievement tests) for all but the highly motivated. The extent to which teachers monitored, encouraged, prodded, etc., their students may well be crucial but this was not tapped by our data collection instruments. Anecdotal experiences gained over the years of association with schools participating in Project LONGSTEP suggest that some students tend to lose momentum unless the teacher provides the kind of individual attention that is unlikely except when a sufficient number of qualified assistants are available.

This problem of student motivation was discussed by Lipson (1974) in his explanation of the possible reasons why the IPI mathematics modules did not produce the dramatic gains that had been expected. To a large extent, his comments apply to a number of other individualized programs. According to Lipson, the process of each student working alone on his instructional module can have a very detrimental effect on motivation, in contrast to a conventional classroom where the group provides competition and a feeling of importance to the activities. "The student's relationship to the other members of the class and to the teacher creates a sense of obligation for some level of performance. The rhythm of the class,

the rituals of class activity--even when we grumble--give our lives a pattern to follow. The progress of the class is like a tide which carries all the swimmers along even though some are slower and some are faster" (1974, p. 60). In concluding, Lipson expresses the hope that new systems will be developed "which build upon the combined strengths of the classroom approach and the modularized and individualized approaches."

The fact that our Degree of Individualization index failed to discriminate between successful and unsuccessful educational approaches also suggests the possibility that we may have overlooked some elements of individualization that occur in conventional classrooms. It could be that individualization is one of the practices that distinguishes good teachers from poor ones, even in traditional educational environments where there is no "experimental" or "innovative" program claiming to be individualized. Perhaps the good teacher individualizes informally and so instinctively that he or she is not even aware of it in these terms. The good teacher does not treat all children as interchangeable cogs; but rather as the individuals they are. This kind of informal individualization would not be noted as a feature of an educational program in Project LONGSTEP--but it may nevertheless be very effective.

This hypothesis, incidentally, if correct, would extend to the elementary school level the interpretation of some loosely comparable findings at the high school level, based on Project TALENT data (Shaycoft, 1967). Analyzing Project TALENT retest data, Shaycoft found evidence, as has Project LONGSTEP, that schools differ in effectiveness. However, she failed to find, as LONGSTEP has failed to find, that these differences were meaningfully related to available variables in any way that might have been expected.

On the other hand, it is also possible that large overall, educational effects were not demonstrated in Project LONGSTEP (or have not been found in many educational evaluation efforts) because such effects have been attenuated by inappropriate matches between educational approaches and student needs. The undoubtedly dramatic growth in achievement demonstrated by a number of students participating in Project LONGSTEP certainly suggests that some near-optimal match of student and educational approach may have been one of the reasons for the gains of these students.

In final summary, the Project LONGSTEP findings should not be viewed as a sweeping criticism of the concepts of innovation or individualization. However, they should serve as a reminder to educators--as well as to parents and legislators--that educational innovation *per se* will not necessarily produce dramatic effects on student achievement. Educational quality is *not* synonymous with innovation or individualization. Individualization of instruction may represent a valuable approach for the improvement of American education, but the findings here suggest that individualization as a program strategy should not be viewed as the final or complete answer. In this case, more (innovation) is not necessarily better (education).

References

Coles, G. J., Chalupsky, A. B., Everett, B. E., Shaycoft, M. F., Rodabaugh, B. J., & Danoff, M. N. Impact of educational innovation on student performance: Project methods and findings for three cohorts. Volume I: Project LONGSTEP final report. Palo Alto, Calif.: American Institutes for Research, 1976.

Lipson, J. I. IPI math--An example of what's right and wrong with individualized modular programs. Learning, March 1974, 2(7), 60-61.

Shaycoft, M. F. The high school years: Growth in cognitive skills. (Interim report 3 to the U. S. Office of Education, Cooperative Research Project No. 3051). Pittsburgh: American Institutes for Research and University of Pittsburgh, 1967.